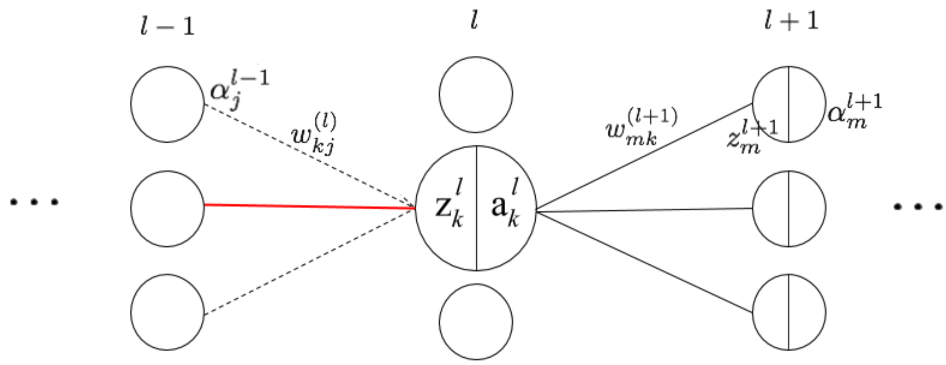

We will prove that the derivative of the cost function with respect to the weight $w_{kl}^{(l)}$ of the l^{th} hidden layer is $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta w_{kl}^{(l)}} = \alpha_j^{l-1} f'(z_k^l) \sum_m \delta_m^{(l+1)} w_{mk}^{l+1}$, where f is the activation function and $\delta_m^{(l+1)}$ is the error propagated from layer $l+1$.

In the following, we will assume zero bias term for the sake of simplicity.

$z_k^l = \sum_j w_{kl}^{(l)} \alpha_j^{l-1}$

$\alpha_k^l = f(z_k^l)$

$z_m^{l+1} = \sum_k w_{mk}^{(l+1)} \alpha_k^l$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta w_{kl}^{(l)}} = \underbrace{\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta z_k^l}}_{\delta_k^l} \cdot \frac{\vartheta z_k^l}{\vartheta w_{kl}^{(l)}} = \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta \alpha_k^l} \cdot \underbrace{\frac{\vartheta \alpha_k^l}{\vartheta z_k^l}}_{f'(z_k^l)} \cdot \underbrace{\frac{\vartheta z_k^l}{\vartheta w_{kl}^{(l)}}}_{\alpha_j^{l-1}} = \left(\sum_m \underbrace{\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta z_m^{l+1}}}_{\delta_m^{l+1}} \frac{\vartheta z_m^{l+1}}{\vartheta \alpha_k^l} \right) \cdot f'(z_k^l) \cdot \alpha_j^{l-1} =$$
$$\left(\sum_m \delta_m^{l+1} w_{mk}^{(l+1)} \right) f'(z_k^l) \cdot \alpha_j^{l-1}$$